

## **REMARKS**

In response to the Office Action dated July 14, 2009, the Assignee (Nuance Communications, Inc.) requests reconsideration. Claims 1-17 were pending. Claims 1-16 are now amended and claim 17 is cancelled. No claims are added. Therefore, claims 1-16 remain pending, with 1, 7, and 13 being independent. No new matter has been added.

### **I. Claim Rejections Under 35 U.S.C. § 101**

Claims 13-17 stand rejected under 35 U.S.C. § 101 as purportedly being directed to non-statutory subject matter because the claims, as previously pending, purportedly could be interpreted as claiming only software. Without conceding to the appropriateness of this rejection, the Assignee has amended claim 13 to recite in part “a text-to-speech engine on a computer...,” thereby addressing the issue raised in the Office Action. Accordingly, withdrawal of the rejections under 35 U.S.C. § 101 is requested.

### **II. Overview of Embodiments**

At any given moment, a voice portal may have an active set of grammars that may correspond to commands a user may input at that moment to interact with the voice portal ([0003]). The active set of grammars may include a current grammar for the actual choice to be made by the user, and additional “universal” grammars (e.g., “go back”) that the user could input ([0003]). Figure 1 provides an example.

As shown, a tree 10 of a voice portal includes choices which a user may make at various selection steps. When the user is at the home directory 11, the user may input either “business” 12 or “entertainment” 14. Those two choices (“business” and “entertainment”) are the current grammars for the choice the user makes at the home directory 11 ([0005]). It should be appreciated, therefore, that the current grammar(s) may change from one menu selection step to another (e.g., the current grammars for home directory 11 may differ from the current grammars for selection step 60) ([0005]). There may also be additional active grammars that the user could input at home directory 11, such as “quit” or “go back” ([0005]). As the user progresses through

the selection steps of the tree, additional grammars may be added to the set of active grammars which the user could input and which the voice system may recognize ([0006]).

Conventional voice portals gave rise to the problem that the voice-based application might incorrectly recognize a voice input when two grammars of the voice portal have similar pronunciations ([0002]-[0003]). An example of this scenario can be seen in FIG. 1, which shows that application 34 and menu option 18 are both titled “Directory.” Therefore, if the grammar for selecting menu option 18 is active within the selection choice following menu option 17, the system would have trouble distinguishing between whether the input received from a user was “Directory” 18 or whether it was “Directory” 34 ([0007]).

According to one embodiment of the application, a method is provided for evaluating the quality of voice recognition by a voice portal ([0009]), and may involve testing the ability of the voice portal to recognize a particular grammar from among the set of other grammars that may be active with the particular grammar being tested ([0017]). Referring to FIG. 1 for an example, if the current grammar is “Projects” 30, the active grammars may include “Meetings” 32, “Directory” 34, “Information” 16, “Director” 18, “Sports” 19, “go back,” and “quit.” ([0023]).

As shown in FIG. 3, the method may involve extracting a current grammar from the voice portal at 210 ([0022]), and generating a test input for that grammar (step 220) ([0024]). The test input, which may include a test pattern (e.g., the actual word or term for the current grammar, additional words, terms, or sounds) may then be input to a voice server and analyzed to assess how well the voice portal recognized the test pattern (steps 230 and 240).

A set of statistics may be derived from the analysis and used to assess the quality of recognition (step 240, [0027]). At step 250, a determination may be made whether the quality of recognition is acceptable. If so, the method may end. If not, the current grammar may be modified and then the method repeated to assess whether the quality of recognition of the voice portal is satisfactory for the modified grammar ([0028]).

It should be appreciated that the foregoing discussion of embodiments is provided merely to assist the Examiner in appreciating various aspects of the present invention. However, not all of the description provided above necessarily applies to each of the independent claims pending in the application. Therefore, the Examiner is requested to not rely upon the foregoing summary

in interpreting any of the claims or in determining whether they patentably distinguish over the prior art of record, but rather is requested to rely only upon the language of the claims themselves and the arguments specifically related thereto provided below.

### **III. Rejections Under 35 U.S.C. § 103**

Independent claims 1 and 7 stand rejected under 35 U.S.C. § 103(a) as purportedly being unpatentable over U.S. Patent No. 7,117,153 (“Mahajan”) in view of U.S. Patent Publication No. 2002/0188451 (“Guerra”) and U.S. Patent Publication No. 2002/0049593 (“Shao”)<sup>1</sup>.

Independent claim 13 stands rejected under 35 U.S.C. § 103(a) as purportedly being unpatentable over Mahajan in view of Guerra and Shao and further in view of U.S. Pat. 6,275,797 (Randic). Reconsideration is requested in view of the claim amendments indicated above and the following remarks.

#### *A. Mahajan*

The portions of Mahajan cited by the Office Action relate to training a language model of a speech recognition system in combination with an acoustic model. Mahajan asserts that speech recognition systems use two types of models: (1) an acoustic model and (2) a language model (col. 2, ll. 10-12). The acoustic model converts features of an acoustic input signal into potential sequences of speech units, the speech units being words or subsets of words (col. 1, ll. 12-14). The language model provides probability distributions for various sequences of words that can be formed from the potential sequences identified by the acoustic model (col. 1, ll. 14-17).

According to Mahajan, a problem existed in that language models were trained without examining how the acoustic model would perform on the language model corpora (col. 1, ll. 39-41). One reason for this was apparently the expense involved in producing acoustic data to examine the performance of the acoustic model (col. 1, ll. 37-39). Mahajan addresses this

---

<sup>1</sup> The Assignee notes that the rejection of claims 1 and 7 on page 3 of the Office Action only mentions Mahajan and Guerra, and not Shao. However, page 5 of the Office Action mentions using Shao in Mahajan and Guerra to purportedly meet the claimed invention. Therefore, the Assignee assumes the statement of the rejection of claims 1 and 7 on page 3 of the Office Action intended to include Shao in the basis for rejection.

concern by developing a so-called confusion model that allows the acoustic model to be examined using text input rather than acoustic input (col. 1, ll. 51-58).

FIG. 2 illustrates a method for constructing and using the confusion model (col. 1, ll. 65-67). The method begins by training an acoustic model (col. 4, ll. 58-60). Once the acoustic model is trained, a confusion model is developed. This involves decoding training data (i.e., an acoustic signal used to train the acoustic model) having a known speech unit sequence using the acoustic model to generate a predicted speech unit sequence (col. 5, ll. 11-13). The predicted speech unit sequence is then aligned to the actual speech unit sequence (step 204). The confusion model can then be constructed from the predicted sequence and the actual sequence (step 206).

The confusion model constructed at 206 can be used to model the performance of the acoustic model of the speech recognition system without the need for an acoustic signal, but rather using a text signal (col. 7, ll. 49-52), thereby impliedly addressing the purported problem mentioned above of the difficulty in producing acoustic data to test the performance of the acoustic model.

At step 208, test text (rather than an acoustic signal) is then decoded using the confusion model from step 206 to produce a predicted speech unit sequence, and probabilities that the predicted speech unit sequence is the actual speech unit sequence (for the test text) are generated. At step 210, errors between the predicted speech unit sequence and the test text are then identified. The method finishes at step 212 with the generation of a word error rate score for the test text of step 208 (col. 9, ll. 37-38). The language model of the voice recognition system can then be trained with the acoustic model using training based on the word error rates calculated from step 21, and may be modified to improve the word error rate score (col. 10, ll. 28-31).

B. Guerra

Guerra relates to dynamically configurable voice portals (see Title). A user inputs utterances to a speech recognition portal, which are interpreted using a speech recognition process (Abstract). Then, one or more aspects of the speech recognition process are dynamically configured (Abstract).

C. Shao

Shao provides a speech processing apparatus and method (see Title). Referring to FIG. 2, a speech recognition system includes a preprocessor 15, a buffer 16, and a recognition block 17 ([0017]). The preprocessor 15 receives electrical signals from microphone 7 representative of input speech, which are then converted to a sequence of parameter frames that are passed to the recognition block 17 via the buffer 16 ([0031]). The recognition block also receives inputs from a word model 19, a language model 21, and a noise model 23 ([0031]-[0032]). The recognition block 17 then outputs the recognized word sequence and a confidence score of how confident the recognition block is that the recognized word sequence accurately reflects the input speech provided to the microphone 7 ([0034]). As shown in FIG. 4, the confidence score is based in part on a best match score and an ambiguity ratio.

D. Randic

Randic relates to testing the quality of a voice path in a communication network using speech recognition (see Abstract). Voice signals are transmitted over a voice path from a sending computer to a receiving computer (Abstract). The receiving computer receives the voice signals, which are then interpreted by a speech recognition engine (Abstract). The speech patterns in the voice signal identified by the speech recognition engine are compared to the reference speech patterns of the voice signal (Abstract). The quality of the voice path over which the voice signal was transmitted can then be determined by comparing the received speech patterns to the reference speech patterns (Abstract).

E. Independent Claim 1 Would Not Have Been Obvious In View Of the Combination of Mahajan, Guerra, and Shao

The proposed combination of Mahajan, Guerra, and Shao would have failed to meet all of the limitations of claim 1, such that claim 1 would not have been obvious in view of the combination. As amended, claim 1 recites:

A method of evaluating grammars associated with a voice portal on a portal server, said method comprising:

- generating a test input for a current grammar of the voice portal, the test input including a test pattern;
- providing the test input to the voice portal on the portal server using a voice server;
- receiving at least one measure of how distinguishable the current grammar is from other grammars of a set of active grammars that are active when the current grammar is active based at least in part on analysis of the test pattern with respect to the set of active grammars, the current grammar being one grammar of the set of active grammars; and

*determining whether to modify the current grammar based at least in part on the at least one measure* (emphasis added).

The combination of Mahajan, Guerra, and Shao would have failed to result in a method meeting at least the above-highlighted limitation of claim 1. There is no discussion in Mahajan, Guerra, or Shao of modifying grammars at all, let alone doing so based on how distinguishable the current grammar is from other grammars of the active grammar. As explained above in section III(A), Mahajan describes modifying a language model to improve a word error rate derived from use of a confusion model. However, Mahajan's language model is not a grammar, and there is no other discussion in Mahajan of modifying grammars. Guerra and Shao also fail to discuss modifying grammars. It should therefore be appreciated that the combination of Mahajan, Guerra, and Shao does not teach or suggest modifying a grammar.

Because the combination does not teach or suggest modifying grammars, it necessarily fails to teach determining whether to modify a grammar in any way. Accordingly, it should be appreciated that the combination fails to teach or suggest "determining whether to modify the current grammar..." as required by claim 1, let alone "determining whether to modify the current grammar based at least in part on the at least one measure." Accordingly, the combination would have failed to meet all of the limitations of claim 1.

For at least this reason, claim 1 would not have been obvious in view of the combination of Mahajan, Guerra, and Shao, and the Assignee requests that the rejection of claim 1 be

withdrawn. Withdrawal of the rejections of claim 2-6 is also requested, since these depend from claim 1 and are patentable for at least the same reasons.

F. *Independent Claim 7 Would Not Have Been Obvious In View of the Combination of Mahajan, Guerra, and Shao*

From the foregoing discussion in section III(E), it should be appreciated that the proposed combination of Mahajan, Guerra, and Shao would have failed to meet all of the limitations of claim 7, such that claim would not have been obvious in view of the combination. For example, claim 7 recites:

A computer-readable storage medium encoded with instructions which, when executed by a computer, cause the computer to perform a method of evaluating grammars associated with a voice portal, the method comprising:  
generating a test input for a current grammar of the voice portal, the test input including a test pattern;  
providing the test input to the voice portal;  
receiving at least one measure of how distinguishable the current grammar is from other grammars of a set of active grammars that are active when the current grammar is active based at least in part on analysis of the test pattern with respect to the set of active grammars, the current grammar being one grammar of the set of active grammars; and  
*determining whether to modify the current grammar based at least in part on the at least one measure* (emphasis added).

The combination of Majahan, Guerra, and Shao would have failed to meet at least the above-highlighted limitation of claim 7, such that the withdrawal of the rejection of claim 7 is requested. Withdrawal of the rejections of claim 8-12 is also requested, since these depend from claim 7 and are patentable for at least the same reasons.

G. *Independent Claim 13 Would Not Have Been Obvious In View of the Combination of Mahajan, Guerra, Shao, and Randic*

The combination of Mahajan, Guerra, Shao, and Randic would not have met all of the limitations of claim 13, such that claim 13 would not have been obvious in view of the combination. For example, claim 13 recites:

A system for evaluating grammars of a voice portal executing on a portal server, the system comprising:  
*an analysis interface for extracting a current grammar from a set of active grammars of the voice portal*, the current grammar being one grammar of the set of active grammars;  
a test pattern generator for generating a test input for the current grammar, the test input including a test pattern;  
a text-to-speech engine on a computer for entering the test input into the voice portal;  
a results collector for analyzing the test input entered into the voice portal against the set of active grammars; and  
a results analyzer for deriving a set of statistics indicative of how distinguishable the current grammar is from other grammars of the set of active grammars (emphasis added).

The combination would have failed to result in a system meeting at least the above-highlighted limitation. Contrary to the assertion in the Office Action, Mahajan does not disclose the claimed “analysis interface.” The Office Action cites to item 304 in FIG. 3 of Mahajan, and to column 5, line 11 of Mahajan (see p. 16 of Office Action referring to rejection of previous claim 17). The cited portion of Mahajan reads:

“At step 202, a portion of training data 304 is spoken by a person 308 to generate a test signal that is decoded using the trained acoustic model.”

That statement fails to say anything about “an analysis interface for extracting a current grammar from a set of active grammars of the voice portal...” It simply says that a speaker speaks training data. Thus, the cited portion of Mahajan clearly does not show the claimed “analysis interface,” and there is nothing else in Mahajan that shows the claimed analysis interface. Guerra, Shao, and Randic fail to remedy this deficiency of Mahajan, such that the combination of the references would not have resulted in a system comprising the claimed analysis interface.

For at least this reason, withdrawal of the rejection of claim 13 is requested. Withdrawal of the rejections of claims 14-16 is also requested, since these depend from claim 13 and are patentable for at least the same reasons.

**IV. Additional Comments on Dependent Claims**

Since each of the dependent claims depends from a base claim that is believed to be in condition for allowance, the Assignee believes it is unnecessary at this time to argue the allowability of each of the dependent claims individually. The Assignee does not, however, necessarily concur with the interpretation of any dependent claim as set forth in the Office Action, nor does the Assignee concur that the basis for the rejection of any dependent claim is proper. Therefore, the Assignee reserves the right to specifically address the patentability of the dependent claims in the future, if necessary.

**CONCLUSION**

A Notice of Allowance is respectfully requested. The Examiner is requested to call the undersigned at the telephone number listed below if this communication does not place the case in condition for allowance.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, the Assignee hereby requests any necessary extension of time. If there is a fee occasioned by this response, including an extension fee, the Director is hereby authorized to charge any deficiency or credit any overpayment in the fees filed, asserted to be filed or which should have been filed herewith to our Deposit Account No. 23/2825, under Docket No.

N0484.70571US00.

Dated: September 14, 2009

Respectfully submitted,

By   
Thomas A. Franklin  
Registration No.: 63,456  
Richard F. Giunta  
Registration No. 36,149  
WOLF, GREENFIELD & SACKS, P.C.  
Federal Reserve Plaza  
600 Atlantic Avenue  
Boston, Massachusetts 02210-2206  
617.646.8000